# HIGH-THROUGHPUT CRYSTALLOGRAPHY FOR LEAD DISCOVERY IN DRUG DESIGN

*Tom L. Blundell**‡, Harren Jhoti‡ and Chris Abell‡§*

Knowledge of the three-dimensional structures of protein targets now emerging from genomic data has the potential to accelerate drug discovery greatly. X-ray crystallography is the most widely used technique for protein structure determination, but technical challenges and time constraints have traditionally limited its use primarily to lead optimization. Here, we describe how significant advances in process automation and informatics have aided the development of high-throughput X-ray crystallography, and discuss the use of this technique for structure-based lead discovery.

**Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1GA, UK.*
*‡Astex Technology, 250 Cambridge Science Park, Milton Road, Cambridge CB4 0WE, UK.*
*§University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, UK.*
*Correspondence to T.L.B.*
*e-mail: tom@cryst.bioc.cam.ac.uk*
DOI: 10.1038/nrd706

Rapid and revolutionary developments in genome sciences, combinatorial chemistry, informatics and robotics are having major impacts on drug discovery. Genome sequencing projects in man and micro-organisms have provided an unprecedented number of potential drug targets. These have given impetus to the study of protein expression (proteomics) and structure (structural genomics[1]), and have allowed a clearer description of drug targets as molecular components of disease processes[2]. At the same time, there is a rapidly expanding range of screening technologies, as well as consolidations in medicinal chemistry arising from the combinatorial approaches that were pioneered in the 1990s. These developments have created an environment for the emergence of new strategies for drug discovery.

In this review, we focus on the use of high-throughput crystallography for structure-based lead discovery — a strategy that combines features of random screening and rational structure-based design. We describe the background to this approach and discuss the underpinning advances in molecular biology, biochemistry, crystallography, chemoinformatics and bioinformatics.

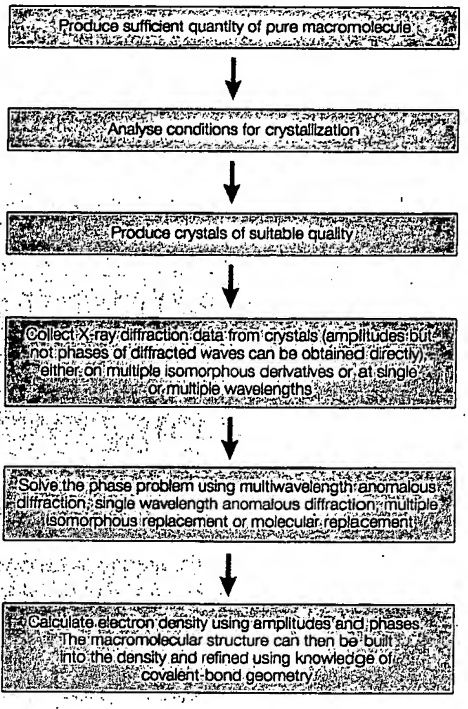
## Background

Two techniques — X-ray crystallography (BOX 1) and nuclear magnetic resonance (NMR) — are used at present for protein structure determination at the atomic level. X-ray crystallography has proved a very versatile method, with most globular macromolecules proving to be crystallizable, and with no limitations on the size and complexity of the macromolecules or their assemblies. NMR has the advantage of being carried out in concentrated solutions rather than in crystals. Comparative studies using the two methods can identify places where crystal contacts disturb the local structure. NMR can define certain dynamic properties of the macromolecules, but it is effectively limited to macromolecules with molecular weights of less than 30 kDa.

Knowledge of the three-dimensional structures of target proteins provides a starting point for structure-based approaches to drug design by defining the topographies of the complementary surfaces of ligands and their protein targets[3,4]. This information can help the synthetic chemist to optimize compounds by building better interactions with the protein, resulting in improved potency and selectivity[5]. Indeed, there are now several drugs on the market that originated from this structure-based design approach. The most commonly cited are human immunodeficiency virus (HIV) drugs, such as amprenavir (Agenerase) and nelfinavir (Viracept), which were developed using the crystal structure of HIV protease[5] (FIG. 1a).

HIV protease was first identified from the HIV genome sequence by the Asp-Thr/Ser-Gly sequence motif of the active site[6], and this was supported by comparative modelling of the three-dimensional structure of the

Box 1 | **Macromolecular crystallography**




An overview of the X-ray crystallography process is shown in the figure. Most globular macromolecules can be ordered as three-dimensional crystals, which amplify the scattering of electromagnetic radiation, giving rise to a diffraction pattern. Intense sources of X-rays are required to obtain diffraction patterns from small crystals of macromolecules, and this is usually achieved using a rotating-anode X-ray generator or a synchrotron. In order to minimize the damage due to heating and free radicals moving in the solvent regions, the crystals are usually flash cooled to liquid-nitrogen temperatures and maintained at this temperature during storage and data collection.

The intensities of the diffraction peaks are proportional to the square of the amplitudes of the scattered waves. These are measured and read out quickly using image plates or fast electronic detectors. Data collection usually requires multiple measurements of equivalent reflections (diffracted waves), leading to the collection and measurement of hundreds of thousands of diffraction intensities. However, in order to reconstitute the image of the macromolecule, knowledge is also required of the phases of each of the diffracted waves. These cannot be measured directly.

The phases have traditionally been estimated by replacing water molecules, or sometimes metal cofactors, with ordered, bound heavy metals in order to produce isomorphous crystals. The diffracted intensities of the derivative crystals are compared with those of the native crystal to work out the phases; this is known as isomorphous replacement. This method, however, has been somewhat superseded by the exploitation of anomalous dispersion. This phenomenon results from the anomalous rate that electromagnetic radiation travels in a material if it contains an element with an absorption edge in the vicinity of the energy of the radiation. Multiple wavelength measurements have the effect of simulating isomorphous replacement and allowing the phases to be calculated; this is called the multiwavelength anomalous diffraction (MAD) method. In some cases, phasing that is based on anomalous diffraction using a single wavelength (SAD) is sufficient for structure determination. The 'anomalous scatterers' are usually selenomethionines that are substituted for methionines in a recombinant protein, but could also be a metal cofactor, an added metal or even added halide or argon atoms.

Combining the estimates of the amplitudes and phases of each reflection using a Fourier synthesis allows calculation of the electron density. The macromolecular structure can usually be built into the density using knowledge of the covalent geometry.

If small molecules are bound to the macromolecule that displace the solvent molecules, but do not disturb the intermolecular packing, the crystals of such a complex will be isomorphous with the parent. Under such circumstances, the position of the bound molecule and the conformational changes in the macromolecule can be established by using difference Fouriers. These provide an image of the difference between the complex and the parent with the density at about half the correct density, as the phases are not quite correct.

dimer on the basis of the structures of aspartic proteases, such as pepsin and renin[7,8]. Homology with renin, already a target in the design of anti-hypertensives, indicated a possible approach to the development of useful inhibitors; similar chemistry could be exploited. When further genome sequences became available, this experience encouraged the use of other three-dimensional structures to aid fold recognition or, more correctly, sequence–structure homology recognition, in order to identify new distant members of the same superfamily that might also be useful targets for drug discovery. More recently, the pioneering development of the flu drug zanamivir (Relenza) involved extensive modelling based on the crystal structure of neuraminidase[9] (FIG. 1b), and resistance problems to the first protein-kinase drug, imatinib (Gleevec), have been rationalized by reference to the crystal structure of the kinase domain of c-ABL[10].

Although pharmaceutical companies continued to embrace structure-based design in the 1990s, as shown by the expansion of in-house crystallographic facilities, the focus of discovery moved to 'diversity-based' screening. This was facilitated by the advent of high-throughput screening (HTS), and the subsequent emergence of combinatorial chemistry. The original visions of combinatorial chemistry invoked large libraries of compounds synthesized on resin beads, which would be screened as mixtures

and then decoded. Many of these concepts have been pragmatically adapted by the pharmaceutical industry. Solid-phase chemistry has proved less tractable than hoped, and the effort of decoding mixtures means that, although many compounds are made, most are screened as single compounds. However, combinatorial chemistry has had a profound effect on the practice and perception of chemistry in the drug discovery process. The number of compounds in company collections has soared by an order of magnitude in the past 10 years, although the quantity and characterization of each compound might be reduced. The level of automation in synthesis has become widespread to such an extent that this is sometimes a determinant in library design.

Initially, combinatorial chemistry offered the panacea of universal libraries. However, the size of chemical space is too large for this to be a realizable or useful concept. Library development then went through a chemistry-driven phase, in which amenable templates (such as purines[11] or shikimic acid[12]) or reactions (such as the Ugi reaction[13]) were exploited, or natural-product skeletons were decorated[14]. The use of such libraries has had less impact on the number of new chemical entities discovered than was originally hoped[15], and there is now a growing emphasis on more rational approaches; for example, the use of 'knowledge-based' or 'focused' screening.

High-throughput screening (HTS) is a key part of the present approach to lead discovery in all of the main pharmaceutical companies. Indeed, HTS is generally the first act in the prosecution of a new target. Many of the assays rely on radioactivity (for example, the scintillation proximity assay) or fluorescence-based approaches. These assays are typically performed in 384-well plates in 20-μl volumes, with the scale mitigating the cost of doing so many assays. The assays aim to identify compounds with $IC_{50}$s lower than 10 μM. Needless to say, it is crucial that such assays are well designed, and also that they do not identify too many false positives.

New paradigms in screening are emerging; for example, there is a growing interest in applying biophysical techniques to lead discovery. Applications of mass spectrometry[16], surface plasmon resonance (SPR)[17], NMR spectroscopy[18,19], single molecule fluorescence[20] and X-ray crystallography[21] to lead discovery have recently been described. There is also a growing recognition that useful information can be obtained from the binding of relatively low-affinity compounds. Such compounds, which would be missed by the HTS campaigns being carried out in most pharmaceutical companies, could provide chemically tractable starting points or identify new binding motifs. This idea is explored in more detail later.

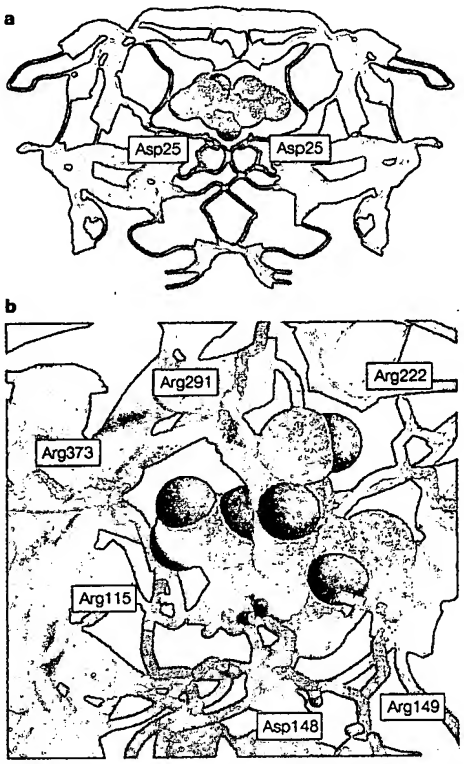



Figure 1 | **Examples of structure-based drug design.** **a** | HIV protease with the inhibitor amprenavir (Agenerase) bound; derived from the crystal structure[75]. The protein is represented with ribbons and amprenavir is shown as a space-filling model with CPK COLOURING. The active-site aspartate residues are highlighted. **b** | Close-up of zanamivir (Relenza) bound to influenza neuraminidase; derived from the crystal structure[10]. Selected residues involved in ligand binding are highlighted.

### High-throughput crystallography

The number of three-dimensional protein structures increased linearly for about 30 years[22], but new technical developments have recently led to an exponential increase in the number of protein structures, similar to that in the number of protein sequences over the previous decade. There are now more than 15,000 three-dimensional structures in the Protein Data Bank (PDB)[22], although these include only ~5,000 different wild-type proteins, the others being duplicates, single mutants, or enzyme–ligand complexes.

There is now intense interest in automating all steps in the protein crystallographic process (BOX 1). Research is being funded by national initiatives, particularly in the United States, Germany and Japan[23]. This has been encouraged by technology drivers, in particular the very intense beams of X-radiation available at many SYNCHROTRON sources, as well as the pull of structural genomics and the use of crystallography in lead discovery and optimization.

Expression, purification and characterization of the proteins in a quantity and form that are suitable for crystallization and X-ray analysis probably occupies over 80% of the time in most structural biology groups. The objective is to obtain about 10 mg, or, if this is not possible, at least 1 mg, of protein that runs as a single band on a denaturing gel and hopefully also on native GEL ELECTROPHORESIS. The protein needs to be soluble at 10 mg $ml^{-1}$ for a good chance of obtaining crystals. The first crucial stage is a thorough analysis of the protein sequence in order to define structural domains that

CPK COLOURING
The CPK colour scheme for elements is based on the colours of the popular plastic space-filling models developed by Corey, Pauling and Kultun, and is conventionally used by chemists. In this scheme, carbon is represented in light grey, oxygen in red, nitrogen in blue and sulphur in yellow.

SYNCHROTRON
A synchrotron accelerates charged particles in a circular orbit. This produces very intense X-rays, which allows the use of smaller and more easily obtained crystals than can be used with conventional X-ray crystallography, and also boosts relevant signals while minimizing noise. The wavelength of synchrotron X-radiation can be varied to perform multiwavelength anomalous diffraction (MAD) experiments.

GEL ELECTROPHORESIS
A method that separates macromolecules on the basis of size, electric charge and other physical properties.

might be suitable for expression. The objective is to define a region that is able to fold on its own and is soluble. It should be free of low-complexity sequences, often found in linker regions, which are likely to have no single conformation, and might interfere with the crystallization. Long loops might also need to be removed if they are non-functional and especially if they are flexible and unstructured; they can often be identified by mild proteolysis. Single membrane-spanning regions are usually engineered out. In addition, post-translational modification needs to be minimal and as homogeneous as possible. Identification of homologues, even if they are distant members of the same superfamily, and construction of a rough comparative model is helpful[24]; this process has now been automated by several groups. A model also provides a basis for systematic mutagenesis of protein surface residues, particularly lysines and glutamates[25], if initial crystallization is unsuccessful.

Methods for high-throughput parallel expression and purification have been developed in many laboratories[26]. *Escherichia coli* expression systems are used if possible, because they are cheaper and quicker than other systems. However, many proteins produced in this way are degraded or produced as insoluble INCLUSION BODIES, so monitoring protein folding before purification avoids time-consuming work. This can be achieved by observing fluorescence of a fusion of the protein to the amino terminus of the GREEN FLUORESCENT PROTEIN (GFP)[27]. Bacterial cell-free systems, such as those based on the *E. coli* S30 extract, offer opportunities for automation of the production of milligram quantities of protein[28]. However, for many disulphide-rich or post-translationally modified eukaryotic proteins, yeast, insect or mammalian expression systems will be necessary. Gel filtration columns that separate on the basis of molecular weight are not very satisfactory in separating out impurities. The use of tags to allow affinity chromatography is preferred, as it can be used to separate compounds that are otherwise very similar, including some minor degradation products, and many proteins can be crystallized with an attached short tag, such as a sequence of six histidines (a six-His-tag)[29].

Over the years, much attention has been focused on the automation of crystallization. Sampling methods, exploiting knowledge of successful precipitating reagents, buffers and pH, have been widely used to reduce the number of crystallization conditions. These have been exploited on a microscale using hanging or sitting drops in the vapour diffusion method, so reducing the amount of protein required. Video systems offer the possibility of using image recognition techniques to monitor crystallization. Now, a new generation of robots that miniaturize the experiments and expand the multidimensional space that is explored are being developed. These can carry out up to $10^5$ trials per day[30,31]. It has been proposed[32] that intrinsic protein fluorescence of single crystals is correlated with their internal order and that this can be used as a rapid method for assessing the resolution and MOSAICITY of the crystals, and thus their suitability for X-ray analysis.

As BOX 1 indicates, once suitable crystals have been produced, it is necessary to collect and process the X-ray data at several wavelengths or on several derivatives, usually by synchrotron radiation. The phases are then defined by multiwavelength anomalous diffraction (MAD) or multiple isomorphous replacement, and used with estimates of the STRUCTURE-FACTOR AMPLITUDES to calculate an electron density map. High-throughput analysis requires automatic storage and mounting of crystals at liquid-nitrogen temperatures. Cassettes or racks that hold up to 96 crystals at liquid-nitrogen temperature, and transfer of the crystals using a robot, have been described[33,34]. Some robots can mount crystals sequentially while maintaining liquid-nitrogen temperature, automatically align the crystal in the beam, collect complete X-ray data sets, and return the crystals to storage. More brilliant synchrotron sources, improved focusing optics and faster read-out detectors have meant that data collection now takes less than 1 hour per crystal. Coping with this high rate of data collection has been greatly assisted by developments in the standard software packages for protein crystallography[35]. However, more completely integrated software environments are required. For example, BLU-ICE (beam-line unification in an integrated control environment), which was developed at the Stanford Synchrotron Radiation Laboratory, is a distributed control system for crystallographic data collection that allows users anywhere to have full control of the experiment[36].

Phase determination has been revolutionized by the application of synchrotron radiation to single wavelength anomalous diffraction (SAD) and MAD techniques[37], using not only substituted selenomethionines, but also other heavy atoms, argon and even simple halides[38]. The determination of phases has been accelerated by powerful Patterson and direct methods for locating the anomalous scatterers[39,40]. New approaches have provided fully automated phasing[41,42].

Many of the new targets for both structural genomics and crystallography in lead discovery will be homologues of previously defined three-dimensional structures. Determination of these structures using molecular replacement is straightforward if there are few conformational changes and sequence identity is >40%, but is often difficult if there are major conformational changes due to ligand binding, or if the search molecule is more distantly related to the target. In such situations, rapid automated molecular replacement by evolutionary search, which systematically explores different structures and conformations, is proving useful[43].

The final objective is to obtain automatic interpretation of the electron density, particularly as 66% of structures are solved at 2.3 Å resolution or better, at which level atomic and molecular fragments are more easily recognized. Perhaps the most spectacular progress in this respect is with the automated refinement procedure, ARP/wARP, in which free atoms and partial structures are included in a hybrid macromolecular model and these are iteratively modelled into the electron density. This approach constructs the structure initially without defining its chemical nature, in order to produce good density, but does allow a complete interpretation of the structure

INCLUSION BODIES
Protein overexpression often leads to the production of insoluble aggregates of misfolded protein, which are known as inclusion bodies.

GREEN FLUORESCENT PROTEIN
Autofluorescent protein originally identified in the jellyfish *Aequorea victoria*.

MOSAICITY
Measure of the degree of order of a crystal. Lower mosaicity indicates better-ordered crystals and hence better diffraction.

STRUCTURE-FACTOR AMPLITUDES
Structure factors are related to the electron density by a mathematical operation called a Fourier transform. Structure-factor amplitudes are determinable from the measured intensities in an X-ray diffraction experiment, but the phases of the diffracted beams, which are needed to reconstitute the electron density, cannot be determined directly.

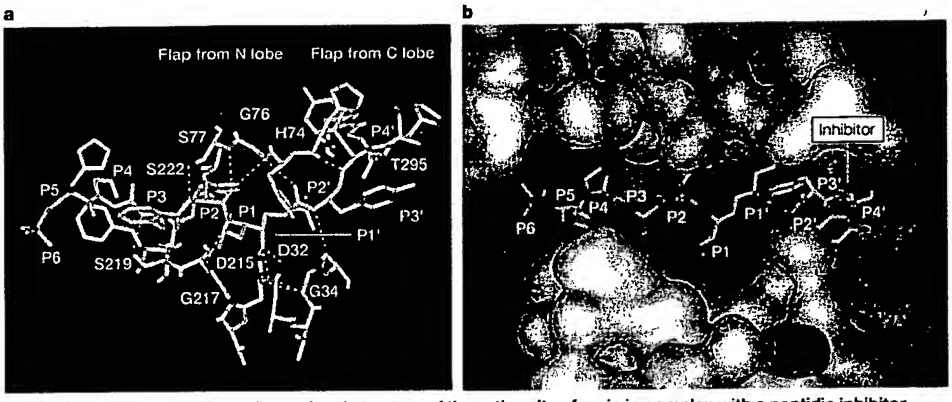



Figure 2 | **Close-up of the three-dimensional structure of the active site of renin in complex with a peptidic inhibitor.** Residues P6–P4′ of the peptidic inhibitor are shown. **a** | The peptide is shown in yellow. Dashed lines indicate hydrogen bonds that are necessary for the inhibitor to have high affinity for renin. Key renin amino-acid residues, including the active-site aspartates (D), are indicated by single letters. The sequence numbering is based on that of pepsin. **b** | Shows the VAN DER WAALS SURFACE of the enzyme, defining the binding pockets for individual side chains that lead to substrate specificity. Residues are coloured according to their electrostatic charge (negative, red; positive, blue). P, peptide residue. (The images were prepared from the structure of mouse renin (Protein Data Bank code 1SMR) by D. Chirgadze using WebLab ViewerPro (Accelrys)).

given a known sequence[44]. It is estimated that this approach is probably already applicable to ~50% of the structures that are being solved at the present time. With new representations of protein structure, this method should soon become applicable to structures of less satisfactory resolution in the region of 3.0 Å. In the not-too-distant future, not only crystallization and data collection, but also complete structure determination, could become routinely possible without human intervention.

**Three-dimensional structures and drug design**

Protein structure determination is now an integral aspect of pharmaceutical research, with most large pharmaceutical companies having both crystallographic and NMR-spectroscopic capabilities. It has also led to the formation of several smaller biotechnology companies that are focusing on high-throughput structure determination. These include Astex Technology Ltd (Cambridge, UK), Integrative Proteomics, Inc. (Toronto, Canada), Plexxikon, Inc. (San Francisco, USA), Proteros Biostructures, GmbH (Martinsreid, Germany), Structural GenomiX, Inc. (San Diego, USA), Syrrx, Inc. (San Diego, USA) and TRIAD Therapeutics (San Diego, USA), all of which have been established recently. Consequently, there is intense competition in obtaining structures of therapeutically important proteins, and such information is vigorously patented.

The accuracy that is required of a macromolecular structure reflects the questions addressed. If the drug designer wishes to know only the general availability of space, essential hydrogen bonds and key electrostatic interactions, a less precise model might be adequate. However, if the design is predicated on the assumption that a lead molecule will precisely complement a known binding site, an accurate model will be required at the highest resolution possible, although computational chemists who are involved in drug design must remember that proteins are flexible and can easily accommodate small changes.

Of course, the accuracy of three-dimensional structures depends on the refinement, the resolution and the restraints that are introduced in the structure analysis. However, much structure-based design seems to assume that the structure is correct, precise and rigid. Modelling software should perhaps oblige the user to know more about the experimental approach, the statistical parameters that indicate the agreement between model and data, and the thermal parameters that give clues about disorder, which are available in the original PDB files.

If there is no three-dimensional structure of the target, a protein with a similar fold can provide the basis for constructing a useful model[45]. For homologous proteins with sequence identities >30%, the common fold can usually be recognized by sequence searches. For more distantly related proteins, profiles or templates are useful in the search for the common fold and alignment of the sequences[46,47]. Once a related fold is identified, this can be used to model the three-dimensional structure. Most methods depend on the assembly of rigid fragments[45], which are used in programs such as COMPOSER to define: first, the framework; second, the structurally variable, mainly loop, regions; and third, the side chains[7]. An alternative approach, encoded in MODELLER[48], seeks to satisfy structural restraints expressed as probability density functions, which are derived from homologues and other proteins. These modelling procedures are most successful if the percentage sequence identity to the unknown target is high (>40%), and obtaining the correct alignment remains an important problem[28,49].

**Interactive graphics and lead development**

Once the three-dimensional structure of a target protein has been defined, it is important to identify the

VAN DER WAALS SURFACE
The van der Waals radius is that which defines the normal contact distance with another non-covalently bound atom. The van der Waals surface is defined by the radii of all such atoms in the molecule.

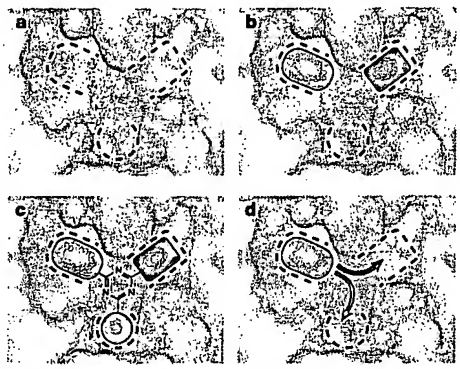



Figure 3 | **Fragment-based screening. a** | A binding site comprising three possible binding pockets. **b** | Crystallographic screening locates molecular fragments that bind into one, two (shown) or all three pockets. **c** | A lead compound is designed by organizing three fragments around a core template or **d** | growing out from a single fragment.

active site and key binding interactions. As many proteins undergo significant conformational changes on ligand binding, the most reliable approach is to determine the structure of a protein–ligand complex, either by co-crystallization or by soaking the ligand into the preformed crystal. The relatively small number of protein–ligand structures in the PDB is in contrast to the wealth of high-resolution small-molecule crystallographic data in the Cambridge Crystallographic Data Centre (CCDC). Furthermore, the protein ligands are reported at much lower resolution than in the CCDC. A cursory inspection of the protein–ligand structures in the PDB, and the data used to derive them, reveals several mistakes in assignment, conformation or description. This is of concern, as it is the primary source of empirical data on protein–ligand interactions. In the absence of the structure of a complex, comparative analysis of ligand binding to homologous structures, site-directed mutagenesis studies, or inactivation studies using agents that are directed at the active site, can be used to deduce the identity of important residues.

Structure-based design begins with the graphical display of hydrogen bonds, molecular surfaces and electrostatic fields. Traditionally, key interactions have been identified visually from three-dimensional structures of macromolecule–ligand complexes. New ligand designs are then explored in an attempt to optimize binding interactions; for example, by optimizing hydrogen bonding and charge–charge interactions (FIG. 2). An important consideration is to minimize the number of rotatable bonds in the ligand to reduce the entropic cost of binding. Increases in affinity can be obtained by introducing hydrophobic groups, although this must not be at the expense of bioavailability. Important general considerations for drug-like molecules, such as solubility and the preferred number of hydrogen-bond donors and acceptors, have been summarized by Lipinski and colleagues[50].

sp³ CARBON
An sp³ carbon has four substituents.

**Docking and virtual ligand screening**

The high cost of experimental binding and screening methods has focused attention on virtual approaches, which are now becoming a useful option if a model of the target protein/receptor and a library of chemical compounds are available[51]. Knowledge of the three-dimensional structure of a drug target allows ligands to be docked into the binding site *in silico*. Docking methods usually precalculate terms for each point on a grid. Goodford[52] pioneered this approach using electrostatic terms for probes, but others developed the method using pseudoenergies that were calculated from pairwise distributions of atoms in protein complexes or crystals of small molecules. This leads to a significant reduction in computational time. In many cases, the correct docking mode can be predicted. A related challenge is to dock a series of probe molecules, fit them to these potentials and rank them according to energy. This is known as virtual ligand screening.

A few methods rely on global optimization of the entire molecule in the receptor field. For example, the Internal Coordinate Mechanics (ICM) docking algorithm[53] uses pseudo-Brownian and torsion moves and a gradient local minimization, and ECEPP3 (the Empirical Conformational Energy Program for Peptides, version 3)[54] uses Monte Carlo minimization. Most methods, however, use fragments. Incremental docking algorithms, such as FlexX[55] and DOCK[49], place fragments in the receptor before constructing the whole ligand. DOCK creates a negative image of the target site and selects and ranks putative ligands on the basis of a comparison of internal distances. Procedures for matching that involve genetic algorithms[56] and graph theory[57] can also be used to generate molecular structures within the constraints of an enzyme active site or a receptor-binding site.

Alternatively, fragments can be positioned in the binding cleft of protein targets and then 'grown' to fill the space available, exploring the electrostatic, van der Waals or hydrogen-bonding interactions that are involved in molecular recognition[58]. For example, GrowMol[59] gives multiple highly diverse structures complementary to active sites, GenStar[60] generates chemically reasonable structures from sp³ CARBONS to fill the binding site, whereas the Multiple Copy Simultaneous Search (MCSS) method[61] maps the binding site out by determining energetically favourable positions and orientations of functional groups on the receptor surface. LUDI[62] positions molecules or new substituents into clefts so that hydrogen bonds are formed and hydrophobic pockets are filled with hydrocarbon groups. Such methods depend on the existence of large databases of small-molecule structures, such as the CCDC, which contains 100,000 crystal structures, or the Fine Chemicals Directory, from which molecular formulae can be automatically processed to a useful three-dimensional representation by CONCORD[63].

Protein flexibility continues to be a challenge. Some methods still use a rigid structure that is based on the structure of the uncomplexed protein, or better on that of the protein complexed to a ligand. Many methods use more permissive or softer models, whereas others take

into account several alternative protein conformations. But methods are now being developed that allow several trial conformations of the protein to be relaxed, and there are still others that cater for joint global optimization of both ligand and protein[55]. Unfortunately, building in flexibility often makes the results worse rather than better, and joint global minimization remains too time-consuming for screening ligands[51].

A further important challenge is the scoring functions. Some scoring functions work by weighting in different physical terms, such as hydrophobicity, solvation electrostatics, hydrogen bonding, ligand deformation energy and van der Waals interaction energy[51]. Alternatively, 'knowledge-based' functions that exploit the statistics of observed inter-atomic contacts can be used, such as the potential of mean force (PMF)[64] and DrugScore[65]. In general, it is found that scoring functions that have been optimized for one protein are an advantage when working with members of the same protein family; for example, a family of homologous proteases. Also, combining different scoring functions seems to be advantageous. A recent comparative analysis[66] showed that GOLD performed rather better than FlexX and DOCK, but the performance depended on the nature of the binding site.

Virtual screening methods can be used to dock a diverse set of drug-like compounds to a protein in minutes, with up to 50% of the molecules docked within 2 Å RMSD (root mean square deviation) of the real structure[51]. From such studies, it is now evident that the task of discriminating a few binders from thousands of non-binders requires special scoring functions. This will depend on distinguishing between correct and incorrect modes of binding.

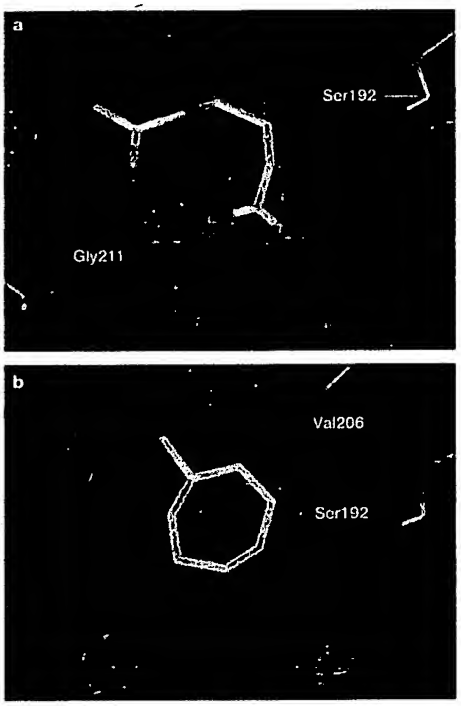



Figure 4 | **Examples of small-molecule fragments bound into a pocket of trypsin. a** | 4-guanidinobutyric acid. **b** | Cycloheptylamine. The electron density was interpreted and models of compounds were automatically fitted using AutoSolve. The electron density maps are contoured at 3σ (1σ is one standard deviation) to ensure that the data are significant, and density due to protein and solvent has been removed for clarity.

**Structure-based lead discovery**

The use of crystallographic and NMR techniques is now being extended beyond structure determination into new approaches for lead discovery. For example, in determining structure–activity relationships (SAR) by NMR[67], perturbations to the NMR spectra of a protein are used to indicate that ligand binding is taking place and to give some indication of the location of the binding site. The ligands can be large molecules, or lower molecular-weight fragments. The experiments are typically carried out using high concentrations of protein (200 μM) and ligand (1–10 mM).

X-ray crystallography has the advantage of defining ligand-binding sites with more certainty. In particular, the binding of small molecules can be studied. This methodology is being developed to identify the molecular fragments that might comprise an inhibitor, and their precise binding interaction with the protein[21,68–70]. The chosen fragments can be soaked (individually or as mixtures) into the target crystals[21,70]. Greer and colleagues[21] describe a method that focuses on soaking the target crystals with cocktails of molecules having differing shapes that can easily be distinguished in the difference electron density, whereas Jhoti and coworkers[70] used automated molecular-fragment matching and fitting to rank candidate fragments in a cocktail, and virtual screening of compounds *in silico* to identify the most suitable molecular fragments[71]. Greer and colleagues have reported the discovery of urokinase inhibitors using a fragment-based approach[21].

The binding of molecular fragments (each with a molecular mass under 200 Da) can potentially give more specific and reliable binding information. For example, if the binding of several aromatic heterocycles is probed against a specific binding pocket in the enzyme (FIG. 3a,b), the discrimination between binding and non-binding will solely be due to the heterocycle–pocket complementarity and will not be modulated by other interactions that might be present in a larger ligand molecule. In this way, new ligand structures that bind to specific protein motifs can rapidly be identified.

In a fragment-based screen, different sets of molecular fragments can be used, analogous to the universal and focused libraries of combinatorial chemistry. For example, in a screen of fragments against trypsin, a focused set was based on benzamidine, 4-aminopyridine and cyclohexylamine, which are known to bind trypsin, and other molecules that are considered capable of making similar interactions, such as histamine, 2-aminoimidazole and 4-aminoimidazole[70]. These molecules were each used as starting points for similarity searches of chemical databases.

PHARMACOPHORE
The ensemble of steric and electronic features that is necessary to ensure optimal interactions with a specific biological target structure and to trigger (or to block) its biological response.


Figure 5 | **Structural screening.** The virtual screening step might involve selection based on chemical similarity, a PHARMACOPHORE and/or large-scale docking into a protein active site. Compounds identified from the virtual screening step are then used in rapid X-ray crystallographic analysis to define their binding modes experimentally. This provides a starting point for iterative chemical elaboration (FIG. 3 c,d) to generate a lead compound.

The molecular fragments are typically dissolved in dimethylsulphoxide (DMSO) and added to a single protein crystal, then left to soak for 1 hour to give the molecule time to penetrate into the protein active site. The concentration of the molecular fragment is typically over 20 mM. This is a much higher concentration than is used in conventional screening experiments, and reflects not only the weakness of the interaction being investigated, but also the high concentration of the protein in the crystal (~10 mM). Compounds can be soaked individually or as mixtures. If mixtures are used, it is best if the individual compounds are unambiguously distinguishable by shape. If too many compounds are soaked at once, solubility can also become a problem. This can be alleviated by using, for example, DMSO as a co-solvent, although the crystal can be damaged if the concentration of DMSO is too high.

As discussed earlier, advances in hardware and software have facilitated high-throughput X-ray crystallography by allowing efficient and speedy collection of data on the soaked crystals. Interpretation and analysis of this data are two key bottlenecks in the process, as there is a need to complex many different compounds to the target (for which the structure is known) and to establish their binding modes rapidly. Conventionally, this requires an experienced X-ray crystallographer to interpret and analyse each X-ray data set collected from a crystal in which the protein has been complexed with a compound, either by co-crystallization or by a soaking experiment. Analysis of a series of isomorphous crystals is less concerned with crystallographic phase determination than with calculation and interpretation of difference Fouriers to position ligands in a previously defined crystal structure (see BOX 1). In order to accelerate this process, it is vital to use automatic procedures, such as AutoSolve[70], which allow the structures of protein–ligand complexes to be solved rapidly by interpreting and analysing the X-ray data without the need for manual intervention. Examples of electron density that were unambiguously interpreted by AutoSolve are shown in FIG. 4; in each case, the binding mode of the small-molecule fragment is clearly defined. It is worth noting that even though the binding affinity of these small-molecule fragments is expected to be in the millimolar range, the binding mode is specific and the key interactions are clearly defined.

If molecular fragments can be found that bind to two independent binding sites, a relatively small library of molecular fragments will sample chemical space very efficiently. Each fragment samples each site, at any given separation and relative orientation. This is a far more comprehensive and elegant screen than having the fragments attached to a rigid template, as might derive from a conventional combinatorial chemistry approach. When the binding of one or more fragments has been determined, this provides a starting point for medicinal chemistry. The fragments can be combined on to a template (FIG. 3c) or be used as the starting point for growing out an inhibitor structure into other pockets on the active site (FIG. 3d).

When all the above processes are coordinated into a seamless process (FIG. 5), they form a rational and powerful approach to lead discovery. Virtual screening coupled with high-throughput X-ray crystallography focuses on identifying one or more weakly binding small-

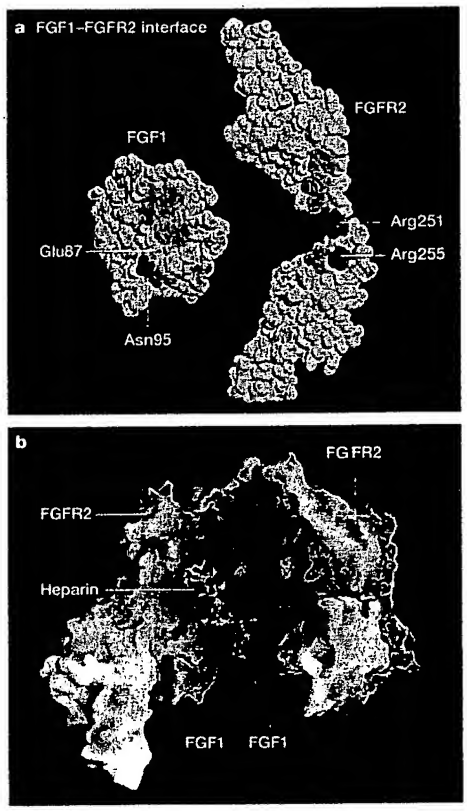



Figure 6 | **Targeting protein–protein interactions. a** | The interaction of the fibroblast growth factor receptor (FGFR) with FGF involves extensive surfaces. **b** | The interaction of heparin, an endogenous small-molecule analogue of heparan sulphate, might provide a better target for drug discovery. (Figure 6b reprinted with permission from *Nature* (REF. 74) © (2000) Macmillan Magazines Ltd.)

molecule fragments from compound libraries that consist of hundreds of small-molecule fragment(s). The high-resolution definition of this binding interaction provides an information-rich starting point for medicinal chemistry. The use of high-throughput X-ray crystallography does not end there, as it becomes a rapid technique to guide the elaboration of the fragments into lead compounds of larger molecular weight.

## Concluding remarks

Pharmaceutical companies have often adopted similar strategies to guarantee that they do not fall behind the competition. This innately conservative approach is now being challenged by the rate of technological development impinging on the industry, the plethora of small biotech companies and the massively complicated patent issues that relate to new targets and screens. It is also being challenged by the paucity of new chemical entities that are emerging from the use of conventional lead-discovery approaches.

It is not long since protein target discovery was an area of intense interest and investment. The problem now, however, is how to manage the numerous targets that are available, and at what stage to consider a target as being validated. It is worth considering that only about 500 targets have been studied[72] in the history of pharmaceutical research, and that there are an estimated 40,000 genes in the human genome, which will correspond to even more proteins in the proteome[73]. Of the enzymes that have been targets so far, almost all have binding sites that are well-defined deep grooves. A small drug candidate can make extensive interactions in these grooves, which compensate for the loss of rotational and translational entropy on binding.

One important challenge for drug discovery arises from the large surfaces that are characteristic of many of the protein complexes, such as those that are involved in receptor recognition and signal transduction. This is illustrated by the interaction of fibroblast growth factor (FGF) with the ectodomain of its receptor tyrosine kinase, FGFR, and with the low-affinity receptor, heparan sulphate[74] (FIG. 6). Not only is it difficult to bind a small molecule to the large, relatively flat surfaces of many proteins involved in protein interactions, it would also be difficult to disrupt the interaction entirely even if one did. It remains to be seen whether the emerging lead-discovery approaches described in this review will prove suitable for these systems.


1. Blundell, T. L. & Mizuguchi, K. Structural genomics: an overview. *Prog. Biophys. Mol. Biol.* **73**, 289–295 (2000).
2. Campbell, S. F. Science, art and drug discovery: a personal perspective. *Clin. Sci.* **99**, 255–260 (2000).
3. Whittle, P. J. & Blundell, T. L. Protein structure-based drug design. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 349–375 (1994).
**A discussion of structure-based lead optimization, describing approaches developed over the previous decade and some of the successes.**
4. Blundell, T. L. Structure-based drug design. *Nature* **384**, S23–S26 (1996).
5. Greer, J., Erickson, J. W., Baldwin, J. J. & Varney, M. D. Application of the three-dimensional structures of protein target molecules in structure-based drug design. *J. Med. Chem.* **37**, 1035–1054 (1994).
6. Toh, H., Ono, M., Saogo, K. & Miyata, T. Retroviral protease-like sequence in the yeast transposon TY1. *Nature* **315**, 691 (1985).
7. Blundell, T. L. *et al.* Knowledge-based protein modelling and design; 18th Sir Hans Krebs Lecture. *Eur. J. Biochem.* **172**, 513–520 (1988).
8. Pearl, L. H., & Taylor, W. R. A structural model for the retroviral proteases. *Nature* **329**, 351–354 (1987).
9. Varghese, J. N. Development of neuraminidase inhibitors as anti-influenza virus drugs. *Drug Dev. Res.* **46**, 176–196 (1999).
10. Schindler, T. *et al.* Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science* **289**, 1938–1942 (2000).
11. Gray, N. S. *et al.* Exploiting chemical libraries, structure, and genomics in the search for kinase inhibitors. *Science* **281**, 533–538 (1998).
**Iterative chemical synthesis and biological screening of 2,6,9-tri-substituted purines are used to develop potent inhibitors of the human CDK2. The structural bases for the binding affinity and selectivity are determined by analyses of the crystal structure of a CDK2–inhibitor complex, and the cellular effects are characterized in yeast by monitoring changes in messenger RNA levels using high-density DNA arrays.**
12. Tan, D. S., Foley, M. A., Shair, M. D. & Schreiber, S. L. Stereoselective synthesis of over two million compounds having structural features both reminiscent of natural products and compatible with miniaturized cell-based assays. *J. Am. Chem. Soc.* **120**, 8565–8566 (1998).
13. Keating, T. A. & Armstrong, R. W. Molecular diversity by a convertible isocyanide in the Ugi 4-component condensation. *J. Am. Chem. Soc.* **117**, 7842–7843 (1995).
14. Nicolaou, K. C. *et al.* Solid and solution phase synthesis and biological evaluation of combinatorial sarcodictyin libraries. *J. Am. Chem. Soc.* **120**, 10814–10826 (1998).
15. Leach, A. R. & Hann, M. M. The *in silico* world of virtual libraries. *Drug Discov. Today* **5**, 326–336 (2000).
16. Moy, F. J. *et al.* MS/NMR: a structure-based approach for discovering protein ligands and for drug design by coupling size exclusion chromatography, mass spectrometry, and nuclear magnetic resonance spectroscopy. *Anal. Chem.* **73**, 571–581 (2001).
17. Myszka, D. G. & Rich, R. L. Implementing surface plasmon resonance biosensors in drug discovery. *Pharm. Sci. Tech. Today* **3**, 310–317 (2000).
18. Hajduk, P. J., Bures, M., Praestgaard, J. & Fesik, S. W. Privileged molecules for protein binding identified from NMR-based screening. *J. Med. Chem.* **43**, 3443–3447 (2000).
19. Feizo, J. *et al.* The SHAPES strategy: an NMR-based approach for lead generation in drug discovery. *Chem. Biol.* **6**, 755–769 (1999).
20. Rigler, R. Fluorescence correlations, single molecule detection and large number screening — applications in biotechnology. *J. Biotechnol.* **41**, 177–186 (1995).
21. Nienaber, V. L. *et al.* Discovering novel ligands for macromolecules using X-ray crystallographic screening. *Nature Biotechnol.* **18**, 1105–1108 (2000).
**Screening techniques that are driven by X-ray crystallography are able to combine lead identification, structural assessment and lead optimization. A method is described that is rapid, efficient and high throughput, and which results in detailed crystallographic structural information. The utility of the method is shown by the discovery and optimization of a new class of urokinase inhibitors for the treatment of cancer.**
22. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
23. Heinemann, U., Illing, G., & Oschkinat, H. High throughput three-dimensional protein structure determination. *Curr. Opin. Biotechnol.* **12**, 348–354 (2001).
24. Burke, D. F. *et al.* An iterative structure-assisted approach to sequence alignment and comparative modeling. *Proteins Struct. Funct. Genet.* **3**, 1–6 (1999).
25. Longenecker, K. L., Garrard, S. M., Sheffield, P. J. & Derewenda, Z. S. Protein crystallization by the rational mutagenesis of surface residues: Lys to Ala mutations promote the crystallization of RhoGD1 *Acta Crystallogr. D* **57**, 679–688 (2001).
26. Lesley, S. A. High throughput proteomics: protein expression and purification in the post-genomic world. *Protein Exp. Purif.* **22**, 159–164 (2001).
**The application of high-throughput screening technologies is the most appropriate response to the challenge of parallel expression and purification of large numbers of gene products.**
27. Waldo, G. S., Standish, B. M., Berendzen, J. & Terwilliger, T. C. Rapid protein folding assay using green fluorescent protein. *Nature Biotechnol.* **17**, 691–695 (1999).
28. Kigawa, T. *et al.* Bacterial cell free systems: Based on *E. coli* S30 extract used for NMR $^{13}C$ and $^{15}N$ labelled proteins. Cell free production and stable isotope labelling of milligram quantities of proteins. *FEBS Lett.* **442**, 15–19 (1999).
29. Crowe, J. *et al.* 6xHis-Ni-NTA chromatography as a superior technique in recombinant protein expression/purification. *Methods. Mol. Biol.* **31**, 371–387 (1994).
30. Stevens, R. C. High-throughput protein crystallization. *Curr. Opin. Struct. Biol.* **10**, 558–563 (2000).
**High-throughput crystallization of proteins has been advanced by exploiting techniques developed for the combinatorial chemistry industry, including robust liquid systems for handling and mixing small volumes. These are assisted by the availability of intense synchrotron X-ray sources, with improved beam-line optics, which are suitable for studying micrometre-sized crystals.**
31. Mueller, U. *et al.* Development for automation and miniaturisation of protein crystallisation. *J. Biotechnol.* **85**, 7–14 (2001).
32. Asanov, A. N., McDonald, H. M., Oldham, P. B., Jedrezejas, M. J. & Wilson, W. W. Intrinsic fluorescence as a rapid scoring tool for protein crystals. *J. Cryst. Growth* **232**, 603–609 (2001).
33. Abola, E., Kuhn, P., Earnest, T. & Stevens, R. C. Automation of X-ray crystallization. *Nature Struct. Biol.* **7**, 973–977 (2000).
34. Muchmore, S. W. *et al.* Automated crystal mounting and data collection in protein crystallography. *Structure* **8**, R243–R246 (2000).
35. Lamzin, V. S. & Perrakis, A. Current state of automated crystallographic analysis. *Nature Struct. Biol.* **7**, 978–981 (2000).
**A goal of structural biology is to improve the underlying methodology of high-throughput determination of three-dimensional structures of biological macromolecules. This will be achieved by development, automation and streamlining of the process of X-ray-crystal structure solution.**
36. Kuhn, P. & Sottis, S. M. Macromolecular structure determination in the post genomic era. *Nucl. Instrum. Methods Phys. Res. A* **467**, 1363–1366 (2001).
37. Hendickson, W. A. & Ogata, C. M. Phase determination from multiwavelength anomalous diffraction measurements. *Methods Enzymol.* **276**, 494–523 (1997).
38. Dauter, Z., Li, M., & Wlodawer, A. Practical experience with the use of halides in phasing macromolecular structures: a powerful tool for structural genomics. *Acta Crystallogr. D* **57**, 239–249 (2001).

39. Sheldrick, G. M. Patterson superposition and *ab initio* phasing. *Methods Enzymol.* **276**, 628–641 (1997).
40. Weeks, C. M. & Miller, R. Optimizing shake and bake for proteins. *Acta Crystallogr. D* **55**, 492–500 (1999).
41. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
42. De La Fortelle, E. & Bricogne, G. Maximum likelihood heavy-atom parameter refinement for the MIR and MAD methods. *Methods Enzymol.* **276**, 590–620 (1997).
43. Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. Rapid automated molecular replacement by evolutionary search. *Acta Crystallogr.* **55**, 484–491 (1999).
44. Perrakis, A., Morris, R. & Lamzin, V. S. Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458–463 (1999).
45. Johnson, M. S., Srinivasan, N., Sowdhamini, R. & Blundell, T. L. Knowledge-based protein modeling. *Crit. Rev. Biochem. Mol. Biol.* **29**, 1–70 (1994).
46. Jones, D. T. GenTHREADER: an efficient and reliable protein fold recognition for genomic sequences. *J. Mol. Biol.* **287**, 797–815 (1999).
47. Shi, J., Blundell, T. L. & Mizuguchi, K. FUGUE: Sequence–structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J. Mol. Biol.* **310**, 243–257 (2001).
48. Sali, A. & Blundell, T. L. Comparative modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).
49. Leach, A. R. & Kuntz, I. D. Conformational analysis of flexible ligands in macromolecular receptor sites. *J. Comput. Chem.* **13**, 730–748 (1992).
50. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Dev. Rev.* **23**, 3–25 (1997).
51. Abagyan, R. & Totrov, M. High-throughput docking for lead generation. *Curr. Opin. Chem. Biol.* **5**, 375–382 (2001).
**Recent improvements in flexible ligand-docking technology are leading to a more central role for computational methods in lead discovery. Docking and screening procedures can select small sets of likely candidates from large libraries of either commercially or synthetically available compounds.**
52. Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **28**, 849–857 (1985).
53. Abagyan, R., Totrov, M. & Kuznetsov, D. A. ICM: a new method for structure modelling and design. *J. Comput. Chem.* **15**, 488–506 (1994).
54. Trosset, J. Y. & Scheraga, H. A. Reaching the global minimum in docking simulations: a Monte Carlo energy minimisation approach using Bezier splines. *Proc. Natl Acad. Sci. USA* **95**, 8011–8015 (1998).
55. Schapira, M., Raaka, B. M., Samuels, H. H. & Abagyan, R. Rational discovery of novel nuclear hormone receptor antagonists. *Proc. Natl Acad. Sci. USA* **97**, 1008–1013 (2000).
56. Payne, A. W. R. & Glen, R. C. Molecular recognition using a binary genetic search algorithm. *J. Mol. Graph.* **11**, 74–91 (1993).
57. Lewis, R. A. Automated site-directed drug design: a method for the generation of general three-dimensional molecular graphs. *J. Mol. Graph.* **10**, 131–143 (1992).
58. Cohen, N. C. & Tschinke, N. Generation of new-lead structures in computer-aided drug design. *Prog. Drug Res.* **45**, 205–243 (1995).
59. Bohacek, R. S. & McMartin, C. Multiple highly diverse structures complementary to enzyme binding-sites — results of extensive application of a *de novo* design method incorporating combinatorial growth. *J. Am. Chem. Soc.* **116**, 5560–5565 (1994).
60. Rotstein, S. H. & Murcko, M. A. GenStar: a method for *de novo* drug design. *J. Comp. Aided Molec. Design* **7**, 23–43 (1993).
61. Miranker, A., & Karplus, M. Functionality maps of binding sites: a multiple copy simultaneous search method. *Proteins* **11**, 29–34 (1991).
62. Bohm, H. J. LUDI-ruled-based automatic design of new substituents for enzyme inhibitor leads. *J. Comput. -Aided Mol. Design* **6**, 593–606 (1992).
63. Rusinko, A. Using CONCORD to construct a large database of three-dimensional coordinates from connection tables. *J. Chem. Inf. Comput. Sci.* **29**, 327–333 (1989).
64. Muegge, I., Martin, Y. C., Hajduk, P. J. & Fesik, S. W. Evaluation of PMF scoring in docking weak ligands of the FK506 binding protein. *J. Med. Chem.* **42**, 2498–2503 (1999).
65. Gohlke, H., Hendlich, M. & Klebe, G. Knowledge-based scoring function to predict protein–ligand interactions. *J. Mol. Biol.* **295**, 337–356 (2000).
66. Bissantz, C., Folkers, G. & Rognan, D. Protein-based virtual screening of chemical databases. *J. Med. Chem.* **43**, 4759–4767 (2000).
67. Shuker, S. B., Hajduk, P. J., Meadows, R. P. & Fesik, S. W. Discovery of high affinity ligands for proteins: SAR by NMR. *Science* **274**, 1531–1534 (1996).
68. Stout, T. J., Sage, C. R. & Stroud, R. M. The additivity of substrate fragments in enzyme-ligand binding. *Structure* **6**, 839–848 (1998).
69. Verlinde, C. L. M. J., Kim, H., Bernstein, B. E., Mande, S. C. & Hol, W. G. J. In *Structure-based Drug Design* (ed. Veerapandian, P.) 365–394 (Marcel Dekker, New York, 1997).
70. Blundell, T. L. *et al. High-throughput X-ray Crystallography for Drug Discovery* (ed. Flower, D.) (Royal Soc. Chem., London, in the press).
71. Walters, W. P., Stahl, M. T. & Murcko, M. A. Virtual screening — an overview. *Drug Discov. Today* **3**, 160–178 (1998).
72. Drews, J. Drug discovery: a historical perspective. *Science* **287**, 1960–1964 (2000).
73. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
74. Pellegrini, L., Burke, D. F., von Delft, F., Mulloy, B. & Blundell, T. L. Crystal structure of fibroblast growth factor receptor ectodomain bound to ligand and heparin. *Nature* **407**, 1029–1034 (2000).
75. Kim, E. E. *et al.* Crystal-structure of HIV-1 protease in complex with VX-478, a potent and orally bioavailable inhibitor of the enzyme. *J. Am. Chem. Soc.* **117**, 1181–1182 (1995).



Acknowledgements

We would like to acknowledge the help of A. Cleasby, M. Hartshorn, E. Southern, I. Tickle, A. Scharff, M. Verdonk, J. Yon and N. Wallace in the preparation of this manuscript.


## Online links

**DATABASES**

**The following terms in this article are linked online to:**

**Medscape DrugInfo:**
http://prominl.medscape.com/drugdb/search.asp
amprenavir | imatinib | nelfinavir | zanamivir

**Protein Data Bank:** http://www.rcsb.org/pdb/
c-ABL | FGF1–FGFR2–heparin | HIV protease | neuraminidase | pepsin | renin

**SWISS-PROT:** http://www.expasy.ch/sprot/sprot-top.html/
GFP

**FURTHER INFORMATION**

**Cambridge Crystallographic Data Centre:**
http://www.ccdc.cam.ac.uk/

**Stanford Synchrotron Radiation Laboratory:**
http://www-ssrl.slac.stanford.edu/

**Access to this interactive links box is free online.**